Les Col·leccions de Cromos

Albert Bellatriu Guillen

Grau en Estadística Aplicada

Universitat Autònoma de Barcelona

Tutor: Xavier Bardina Simorra

20 de gener de 2015

Índex

Les Col·leccions de Cromos

Resum

En aquest treball considerarem el problema del Col·leccionista de Cromos: quants cromos hem de comprar per completar una col·lecció? Presentarem diferents mètodes per calcular el temps d'espera per completar-la. Considerarem diferents casos, segons si els cromos arriben d'un en un o en grups de mida constant, independentment i amb probabilitats desiguals.

Resumen

En este trabajo consideraremos el problema del Coleccionista de Cromos: cuantos cromos debemos comprar para completar una colección? Presentaremos diferentes métodos para calcular el tiempo de espera para completarla. Consideraremos diferentes casos, según si los cromos llegan de uno en uno o en grupos de tamaño constante, independientemente y con probabilidades desiguales.

Abstract

In this note we will consider the Coupon Collectors Problem: how many coupons we have to purchase to complete a collection? We will present some results to evaluate the expected waiting time to complete it. We will consider diferents cases, in the case of coupons which arrives one by one or in groups of constant size, independently and with unequal probabilities.

1 Introducció

El problema del col·leccionista de cromos és un problema clàssic en probabilitat combinatòria.

Considerem que es vol completar una col·lecció i assumim que la col·lecció és finita, amb N cromos. Ens referirem a cada cromo amb els valors 1,...,N. Definirem les probabilitats d'obtindre el cromo k per p_k.

Les preguntes que ens plantegem són:

- Quants cromos hem de comprar per completar la col·lecció de N cromos?
- Com canvien els resultats si aquests arriben d'un en un o en grups de mida constant?
- I si les probabilitats d'obtindre cada cromo o grup de cromos són desiguals?

En el cas que els cromos arribin d'un en un el problema es simplifica. També el fet que les probabilitats d'obtindre cada cromo siguin iguals o desiguals afecta a la simplicitat dels resultats.

Veurem diferents maneres de calcular el temps d'espera per completar la col·lecció i en algún cas considerarem una aproximació a la fórmula obtinguda, ja que, tot i que les fórmules semblin senzilles, algunes són computacionalment lentes.

També verificarem que el mínim de cromos esperat per completar una col·lecció es dóna en el cas que les probabilitats d'obtindre cada cromo siguin iguals. Veurem també que en el cas de les probabilitats desiguals, un cromo amb probabilitats ínfimes d'apareixer, pot fer que el temps d'espera per completar la col·lecció tendeixi a infinit.

Història

El problema del col·leccionista de cromos, ha sigut considerat pels probabilístics des del segle XVIII. El problema va aparèixer per primera vegada al 1708, al text "*De Mensura Sortis*"(*On the Measurement of Chance*), escrit per A. De Moivre. A finals del mateix segle, Laplace i Euler van donar proposar altres resultats en el cas de probabilitats constants.

Aquests autors van suposar que les probabilitats d'obtindre cada cromo eren les mateixes, i no va ser fins al 1954 quan H.Von Schelling va resoldre per primera vegada el problema de calcular el temps d'espera per completar una col·lecció considerant que les probabilitats d'obtindre cada cromo eren desiguals.

Al 1960 D.J. Newman i L.Sheep, van contribuir al problema donant un nou punt de vista. Van proposar-se calcular el temps d'espera per completar dues col·leccions en el cas que els cromos apareixessin amb les mateixes probabilitats. Van confirmar que s'havien de comprar menys cromos si es volien completar dues col·leccions en comú enlloc de fer-les per separat.

En els últims anys, M. Ferrante, de la "Università degli Studi di Padova, Italia", ha publicat diferents articles sobre aquest tema recordant les fórmules obtingudes a la dècada dels anys 50 i proposant alternatives computacionalment més senzilles.

Aplicacions

El problema del col·leccionista s'usa principalment en enginyeria elèctrica, on està relacionat amb "*to the cache fault problem*", i serveix per detectar fallades elèctriques.

També s'usa en biologia per estimar el nombre de subespècies d'animals d'una certa família.

2 Arribada unitària amb probabilitats iguals

En aquest capítol estudiarem el nombre esperat de cromos que s'han d'adquirir per completar una col·lecció. Considerem que la col·lecció té N cromos diferents, i que la probabilitat d'obtindre cada cromo és la mateixa, $p_k = \frac{1}{N}$.

Primer de tot presentarem l'exemple més clàssic, aquell que trobem habitualment en llibres de probabilitat. Considerarem també un mètode alternatiu per calcular el temps d'espera, i finalment obtindrem una aproximació que facilitarà els càlculs en el cas de col·leccions grans.

2.1 Càlcul usant la distribució geomètrica

Def. 2.1.1. Diem que una variable aleatòria segueix una llei geomètrica de paràmetre p si és el resultat d'un experiment aleatori que només pot tenir dos resultats, diem èxit i fracàs, amb probabilitats respectives p i $q = (1-p)$.

Si designem per X el nombre de proves independents fins que obtenim un èxit, aleshores:

$$\mathbb{P}\{X = m\} = pq^{(m-1)} = p(1-p)^{(m-1)}.$$

Conseqüentment, l'esperança d'una variable aleatòria que segueix una llei geomètrica de paràmetre p és:

$$\mathbb{E}[X] = \sum_{m=1}^{\infty} mpq^{(m-1)} = \frac{p}{(1-q)^2} = \frac{1}{p}.$$

Considerem el problema del col·leccionista que vol saber quants cromos hauria d'esperar comprar per completar una col·lecció de N cromos diferents, que s'adquireixen d'un en un independentment, i cada cromo té les mateixes probabilitats de sortir.

Designant per X el nombre de cromos que necessitem per completar la col·lecció, podem descompondre $X = X_1 + X_2 + ... + X_N$, on per cada $i \in \{1, 2, ..., N\}$, X_i denota el nombre de cromos a comprar per passar de tenir $(i-1)$ cromos diferents a tenir-ne i diferents a la col·lecció.

Amb aquesta definició, X_i és una variable aleatòria amb llei geomètrica on l'èxit és aconseguir un cromo que encara no tinguem i el fracàs obtindre'n un que ja tenim a la col·lecció.

Quan tenim $(i-1)$ cromos diferents, la probabilitat d'obtindre'n un que encara no tinguem és $\frac{(N-i+1)}{N}$.

Com que hem considerat que els cromos apareixen de manera independent, aleshores la variable aleatòria X_i és independent de X_j per tot $j \neq i, j \in \{1, \ldots, N\}$ i segueix una llei geomètrica amb paràmetre $p_i = \frac{(N-i+1)}{N}$.

El temps d'espera per completar la col·lecció és l'esperança de la variable aleatòria X.

Usant la linealitat de l'esperança podem calcular nombre esperat de cromos a comprar per completar la col·lecció:

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{E}[X_1 + ... + X_N] \\ &= \mathbb{E}[X_1] + \mathbb{E}[X_2] + ... + \mathbb{E}[X_N] \\ &= \frac{1}{p_1} + \frac{1}{p_2} + \ldots + \frac{1}{p_N} \\ &= 1 + \frac{N}{N-1} + \frac{N}{N-2} + ... + \ldots + \frac{N}{2} + N. \end{aligned}$$

Reordenant la sèrie, podem reescriure el resultat del temps d'espera per completar una col·lecció de N cromos que arriben de manera unitària amb probabilitats iguals com:

$$\mathbb{E}[X] = N \sum_{i=1}^{N} \frac{1}{i}. \tag{1}$$

Exemple 2.1.2. Hi ha molts casos reals de col·leccions on obtenim els cromos d'un en un. En aquest exemple considerarem diferents valors de N i calcularem el temps d'espera estimat per completar les col·leccions.

N	5	10	30	50	100	200	500
$\mathbb{E}[\boldsymbol{X}]$	11.42	29.29	119.85	224.96	518.74	1175.60	3396.41

Taula 1: Temps d'espera per completar les col·leccions.

2.2 Càlcul mitjançant Cadenes de Markov

En aquest apartat considerarem el mateix problema, però des d'un altre punt de vista, com un problema de processos estocàstics, i comprovarem que el resultat obtingut és el mateix.

Calcularem el temps d'espera per completar una col·lecció, on les probabilitats d'aconseguir cada cromo són les mateixes, i els adquirim d'un en un, igual que en l'apartat anterior.

Def. 2.2.1. Un procés estocàstic és un conjunt de variables aleatòries X_t, indexades per un paràmetre temporal t que pot ser discret o continu. En el nostre problema, considerem el temps com un paràmetre discret.

Si considerem que cada cromo arriba a una unitat de temps, denotant per X_i el temps d'espera per obtindre l'i-èssim cromo diferent quan ja en tenim $i-1$ diferents, el problema es pot resoldre mitjançant cadenes de Markov a temps discret.

Def. 2.2.2. Una cadena de Markov és una successió de variables aleatòries $\{X_n, n \geq 0\}$ que prenen valors en un conjunt finit E, l'espai d'estats, i que compleixen que $\forall i_0, i_1, \ldots, i_{n+1} \in E$,

$$\mathbb{P}(X_{n+1} = i_{n+1} | X_0 = i_0, \ldots, X_n = i_n) = \mathbb{P}(X_{n+1} = i_{n+1} | X_n = i_n),$$

sempre que $\mathbb{P}(X_0 = i_0, \ldots, X_n = i_n) > 0$.
Aquesta propietat s'anomena propietat de Markov.

Col·loquialment la propietat de Markov diu que conegut el present X_n, el futur X_{n+1} és independent del passat $X_0, \ldots, X_{n-1}$.

Sigui Y_n el nombre de cromos diferents que tenim al comprar-ne n, i suposem que la probilitat d'obtindre cada cromo és $p = \frac{1}{N}$. La variable Y_n és una cadena de Markov a temps discret amb espai d'estats $E = \{0, 1, ..., N\}$ amb $|E| = N + 1$.

Amb aquestes definicions obtenim la matriu de transició (a un pas):

$$P = \begin{pmatrix} 0 & 1 & 0 & \dots & \dots & \dots & 0 \\ 0 & \frac{1}{N} & \frac{N-1}{N} & 0 & \dots & \dots & 0 \\ 0 & 0 & \frac{2}{N} & \frac{N-2}{N} & \dots & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & \dots & 0 & \frac{N-1}{N} & \frac{1}{N} \\ 0 & \dots & \dots & \dots & 0 & 0 & 1 \end{pmatrix}$$

Veiem que {N} és l'única classe recurrent, que només conté l'estat N, i que és absorvent.

Calcular el nombre de cromos per completar la col·lecció és el mateix que calcular el temps que triga la cadena Y_n en arribar a l'estat N, ja que hem suposat que a cada instant de temps comprem un cromo.

Def. 2.2.3. En les cadenes de Markov, el temps d'espera pel primer pas per l'estat j partint de l'estat i és:

$$\mu_{ij} = 1 + \sum_{k \neq j, k \in C} p_{ik}\mu_{kj}, \quad \forall i, j \in C \text{ classe recurrent},$$

on p_{ik} és la probabilitat de passar de l'estat i a l'estat k en un sol pas, i per tant, els p_{ik} són els coeficients de la matriu de transició P.

En el nostre cas relacionem l'estat inicial amb el moment que comencem la col·lecció, i volem calcular el temps d'espera pel primer pas per l'estat N, que será el temps d'espera per completar la col·lecció, i de fet, el nombre de cromos que esperem comprar per completar la col·lecció.

Definim $t_i = \mu_{iN}$ que és el temps d'espera per passar de tenir i cromos diferents a completar la col·lecció, per tant ens interessa el valor de $t_0 = \mu_{0N}$.

Construim el sistema lineal:

$$\begin{cases} t_N = 0 \\ t_i = 1 + \sum_{j \neq N} p_{ij} t_j, \quad i \neq N \end{cases}$$

$$\begin{cases} t_N = 0 \\ t_{N-1} = 1 + \frac{N-1}{N} t_{N-1} \\ t_{N-2} = 1 + \frac{N-2}{N} t_{N-2} + \frac{2}{N} t_{N-1} \\ \vdots \\ t_1 = 1 + \frac{1}{N} t_1 + \frac{N-1}{N} t_2 \\ t_0 = 1 + t_1 \end{cases}$$

$$\begin{cases} t_N = 0 \\ \frac{1}{N} t_{N-1} = 1 \\ \frac{2}{N} t_{N-2} = 1 + \frac{2}{N} t_{N-1} \\ \vdots \\ \frac{N-1}{N} t_1 = 1 + \frac{N-1}{N} t_2 \\ t_0 = 1 + t_1 \end{cases}$$

$$\begin{cases} t_N = 0 \\ t_{N-1} = N \\ t_{N-2} = t_{N-1} + \frac{N}{2} = N + \frac{N}{2} \\ \vdots \\ t_1 = t_2 + \frac{N}{N-1} = N + \frac{N}{2} + \ldots + \frac{N}{N-2} + \frac{N}{N-1} \\ t_0 = t_1 + 1 = N + \frac{N}{2} + \ldots + \frac{N}{N-2} + \frac{N}{N-1} + 1 \end{cases}$$

Finalment, obtenim que el temps d'espera perque la cadena passi de l'estat 0 a l'estat N, és:

$$\mu_{0N} = t_0 = N \sum_{i=1}^{N} \frac{1}{i}. \tag{2}$$

Hem vist que el temps d'espera per completar la col·lecció, calculat amb cadenes de Markov, coincideix amb el resultat que hem obtingut a l'apartat anterior.

Concluim que el temps d'espera per completar una col·lecció de N cromos, que aconseguim de manera unitària i tots els cromos amb la mateixa probabilitat, es pot calcular mitjançant la fórmula:

$$\mathbb{E}[X] = N \sum_{i=1}^{N} \frac{1}{i}.$$

A la figura 1 podem observar com varia el temps d'espera per completar una col·lecció en funció del nombre de cromos diferents que hi hagi.

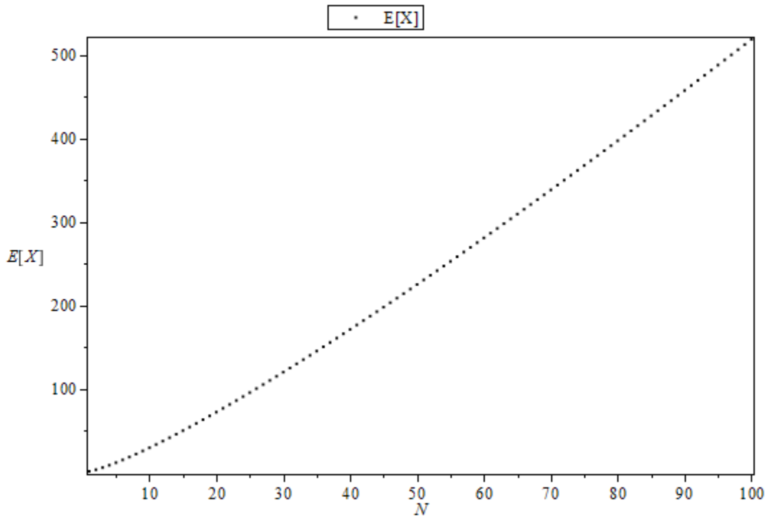

Figura 1: Temps d'espera per completar una col·lecció.

2.3 Una aproximació molt exacta

En els apartats anteriors hem obtingut una fórmula senzilla per resoldre el problema del col·leccionista de cromos, però aquesta inclou una sèrie harmònica en els càlculs, i això fa que sigui lent de calcular per col·leccions grans.

Per obtindre el nombre esperat de cromos a comprar per completar una col·lecció de cromos de mida gran podem utilitzar algún programa de càlcul, com ara Maple, o aproximar la sèrie harmònica inclosa en el resultat obtingut.

La fórmula obtinguda conté la sèrie harmònica

$$\sum_{i=1}^{N} \frac{1}{i} = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{N}.$$

Aproximem aquesta sèrie per

$$\sum_{i=1}^{N} \frac{1}{i} = \log(N) + \gamma + \frac{1}{2N} + O(\frac{1}{N^2}),$$

on $\gamma = \pm 0.5772156649$ és la constant de Euler-Mascheroni.

Podem aproximar $\mathbb{E}[X]$ quan $N \to \infty$ per:

$$\widehat{\mathbb{E}[X]} = N\log(N) + N\gamma + \frac{1}{2} + O(\frac{1}{N}). \tag{3}$$

De fet, podem aproximar el resultat per qualsevol valor de N, ja que, l'aproximació dista del resultat exacte com a màxim 0.1 per qualsevol $N \in \mathbb{N}$.

N	1	2	3	4	5	6	7
$\mathbb{E}[X]$	1	3	5.5	8.33	11.417	14.7	18.15
$\widehat{\mathbb{E}[X]}$	1.077	3.041	5.527	8.354	11.433	14.714	18.162
$\widehat{\mathbb{E}[X]} - \mathbb{E}[X]$	0.077	0.041	0.027	0.021	0.016	0.014	0.012

Taula 2: Error de laproximació $\widehat{\mathbb{E}[X]}$.

Com podem veure a la taula 2 i a la figura 2, la diferència és menor que 0.1 en el cas d'estimar una col·lecció amb un cromo i tendeix a zero ràpidament a mesura que N augmenta. Per tant, podem considerar que es tracta d'una molt bona aproximació.

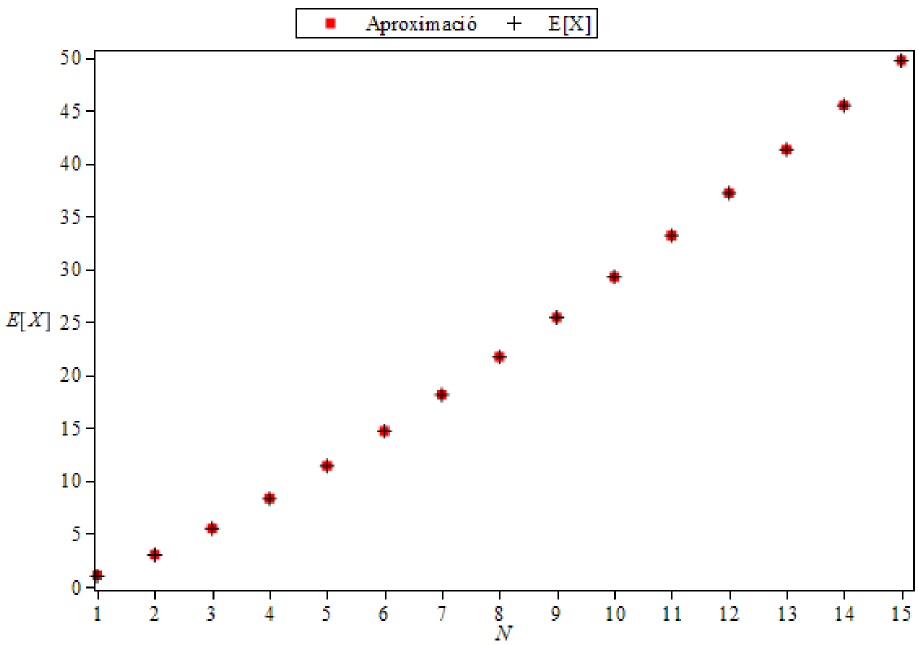

Figura 2: Temps d'espera per completar una col·lecció.

Usant l'aproximació podem calcular fàcilment el temps d'espera per completar una col·lecció de N cromos que arriben de manera unitària i amb les mateixes probabilitats, en el cas que N sigui gran. A la figura 3 podem veure els resultats per col·leccions de fins a mil cromos diferents.

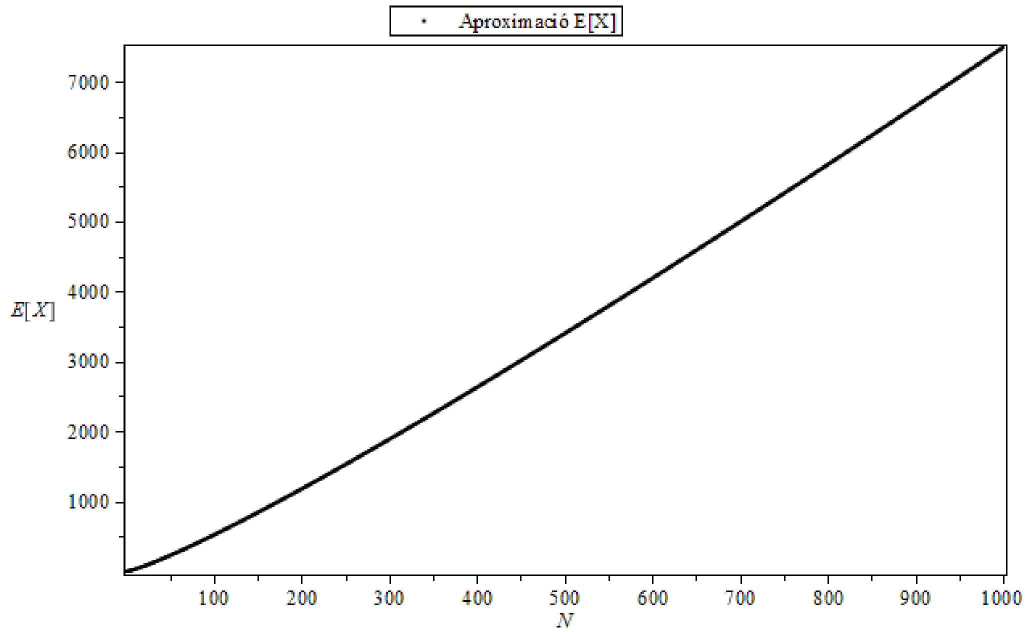

Figura 3: Temps d'espera per completar una col·lecció gran.

3 Arribada unitària amb probabilitats desiguals

En aquest capítol, seguirem calculant el temps d'espera per completar una col·lecció de N cromos, que obtenim de manera individual, però considerarem que cada cromo poT aparèixer amb una probabilitat diferent.

Es tracta d'un cas més real, ja que, les nostres experiències ens fan creure que en tota col·lecció hi ha algun cromo que sembla impossible d'aconseguir, i en canvi n'hi ha d'altres que els acabem tenint repetits moltes vegades.

Més formalment, considerem que la probabilitat d'obtindre el cromo i és $p_i > 0$, amb $p_1 + p_2 + \ldots + p_N = 1$. No contemplem els casos on $p_i = 0$, per algun i, ja que aleshores resultaria impossible completar la col·lecció.

3.1 Càlcul mitjançant Maximums-Minimums Identity

Designem amb X_i el nombre aleatori de cromos que hem de comprar per aconseguir una unitat del cromo i. Per obtindre el nombre de cromos necessaris per completar la col·lecció considerem el màxim de les variables X_i, és a dir, definim X com: $X = \max\{X_1, \ldots, X_N\}$.

L'esperança del nombre de cromos a comprar per completar la col·lecció resulta:

$$\mathbb{E}[X] = \mathbb{E}[\max\{X_1, \ldots, X_N\}] = \mathbb{E}[\max_i X_i]$$

Observació 3.1.1. Les variables X_i segueixen distribucions geomètriques de paràmetre p_i, però en aquest cas no són idènticament distribuides.

El nombre de cromos necessaris per obtindre el primer cromo del tipus i o del tipus j, és $\min(X_i, X_j)$, és una variable aleatòria geomètrica de paràmetre $p_i + p_j$, per $i \neq j$. El mateix succeeix amb qualsevol subconjunt finit de variables aleatòries.

Per calcular l'esperança necessitem la fórmula de "Maximum-minimums Identity". Aquesta fórmula permet canviar el màxim inclòs en el càlcul de l'esperança pel càlcul de mínims, que podrem calcular seguint l'observació anterior.

Observació 3.1.2. Les variables aleatòries X_i són finites amb probabilitat 1, ja que:

$$\begin{aligned}\sum_{n=1}^{+\infty} \mathbb{P}(X_i = n) &= \sum_{n=1}^{+\infty} (1-p_i)^{n-1} p_i \\ &= p_i \frac{1}{1-(1-p_i)} = 1\end{aligned}$$

Proposició 3.1.3. Fórmula "Maximums-Minimums Identity"
Per qualsevol successió de variables aleatòries $X_1, \cdots, X_N$, positives i finites, tenim que:

$$\begin{aligned}\max_i X_i = \sum_i X_i - \sum_{i<j} \min(X_i, X_j) + \sum_{i<j<k} \min(X_i, X_j, X_k) - \dots \\ \dots + (-1)^{(N+1)} \min(X_1, \dots, X_N)\end{aligned}$$

Demostració. Considerem x_i un conjunt arbitrari de valors continguts a l'interval $[0,1]$. Sigui U una variable aleatoria uniforme, $U \in (0,1)$. Considerem els esdeveniments $A_i, i = 1, \cdots, n$ com $A_i = \{U < x_i\}$.

D'aquesta manera tenim que:

$$\cup_i A_i = \{U < \max_i x_i\}$$

Definint aquests esdeveniments obtenim dues igualtats molt útils:

$$\begin{aligned}\mathbb{P}(A_i) = \mathbb{P}\{U < x_i\} = x_i \\ \mathbb{P}(\cup_i A_i) = \mathbb{P}\{U < \max_i x_i\} = \max_i x_i\end{aligned}$$

De la mateixa manera, obtenim igualtats últils amb les interseccions:

$$\begin{aligned}A_{i_1} \cap \dots \cap A_{i_r} = \{U < \min_{j=1,\dots,r} x_{i_j}\} \\ \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_r}) = \mathbb{P}\{U < \min_{j=1,\dots,r} x_{i_j}\} = \min_{j=1,\dots,r} x_{i_j}\end{aligned}$$

La proposició està basada en la fórmula d'inclusió-exclusió per la probabilitat de la unió d'esdeveniments:

$$\begin{aligned}\mathbb{P}(\cup_i A_i) = \sum_i \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \sum_{i<j<k} \mathbb{P}(A_i \cap A_j \cap A_k) - \dots \\ \dots + (-1)^{(n+1)} \mathbb{P}(A_1 \cap \dots \cap A_n)\end{aligned}$$

Usant les igualtats anteriors, podem transformar la fórmula de la unió esdeveniments al càlcul del màxim d'un conjunt de variables aleatòries.

$$\max_i x_i = \sum_i x_i - \sum_{i<j} \min(x_i, x_j) + \sum_{i<j<k} \min(x_i, x_j, x_k) - \dots$$
$$\dots + (-1)^{(N+1)} \min(x_1, \dots, x_N)$$

Hem demostrat la fórmula de "Maximums-Minimums" per qualsevol conjunt de variables aleatòries contingudes a l'interval unitari.

$\square$

En la demostració hem suposat que $x_i \in [0,1]$, però en el nostre cas X_i és una variable aleatòria positiva qualsevol. En el cas que $X_i \in [0,a], \forall i$ amb $a > 1$, considerem la variable transformada $y_i = \frac{X_i}{a} \in [0,1]$. Aleshores, tenim que y_i compleix les propietats citades anteriorment i podem aplicar la fórmula:

$$\max_i y_i = \sum_i \frac{X_i}{a} - \sum_{i<j} \min(\frac{X_i}{a}, \frac{X_j}{a}) + \sum_{i<j<k} \min(\frac{X_i}{a}, \frac{X_j}{a}, \frac{X_k}{a}) - \dots$$
$$\dots + (-1)^{(N+1)} \min(\frac{X_1}{a}, \dots, \frac{X_N}{a})$$
$$\max_i y_i = \frac{\sum_i X_i}{a} - \frac{\sum_{i<j} \min(X_i, X_j)}{a} + \frac{\sum_{i<j<k} \min(X_i, X_j, X_k)}{a} - \dots$$
$$\dots + (-1)^{(N+1)} \frac{\min(X_1, \dots, X_N)}{a}$$

Per simplifcar la lectura definim:

$$M = \sum_i X_i - \sum_{i<j} \min(X_i, X_j) + \sum_{i<j<k} \min(X_i, X_j, X_k) - \dots$$
$$\dots + (-1)^{(N+1)} \min(X_1, \dots, X_N)$$

Hem vist que:

$$\max_i y_i = \frac{M}{a}$$

Per altra banda, hem definit $y_i = \frac{X_i}{a}$, per tant:

$$\max_i y_i = \max_i \frac{X_i}{a} = \frac{1}{a} \max_i X_i$$

Finalment, podem concloure que:

$$\max X_i = M$$

La fórmula de "Maximums-Minimums" serveix per calcular el màxim de qualsevol successió de variables aleatòries positives i finites. De fet, es pot demostrar que la fórmula serveix per calcular el màxim de qualsevol successió de variables aleatòries finites.

Calculem l'esperança del punt 3.1 aplicant la fórmula que acabem de demostrar i obtenim:

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}[\max_i X_i] \\
&= \sum_i \mathbb{E}[X_i] - \sum_{i<j} \mathbb{E}[\min(X_i, X_j)] + \sum_{i<j<k} \mathbb{E}[\min(X_i, X_j, X_k)] - \ldots \\
&\quad \ldots + (-1)^{(N+1)} \mathbb{E}[\min(X_1, \ldots, X_N)]
\end{aligned}$$

Usant la observació 3.1.1, podem calcular les esperances:

$$\mathbb{E}[X] = \sum_i \left(\frac{1}{p_i}\right) - \sum_{i<j} \left(\frac{1}{p_i + p_j}\right) + \sum_{i<j<k} \left(\frac{1}{p_i + p_j + p_k}\right) - \ldots$$

$$\ldots + (-1)^{(N+1)} \left(\frac{1}{p_1 + \ldots + p_N}\right) \tag{4}$$

Hem obtingut el temps d'espera per completar una col·lecció de N cromos, que obtenim de manera individual, i sabent que el cromo i té una probabilitat d'aparèixer de p_i.

Exemple 3.1.4. Veurem casos particulars per calcular el temps d'espera per completar col·leccions amb cromos que arriben amb probabilitats desiguals:

- Per $N = 2, \mathbb{E}[X_{(\frac{1}{2}, \frac{1}{2})}] = 3$

  - Si $p_1 = \frac{1}{3}, p_2 = \frac{2}{3}$, aleshores:

    $\mathbb{E}[X_{(\frac{1}{3}, \frac{2}{3})}] = \frac{1}{p_1} + \frac{1}{p_2} - \frac{1}{p_1 + p_2} = 3 + \frac{3}{2} - 1 = \frac{7}{2} = 3.5$

- ◦ Si $p_1 = \frac{1}{10}, p_2 = \frac{9}{10}$, aleshores:

  $$\mathbb{E}[X_{(\frac{1}{10},\frac{9}{10})}] = \frac{1}{p_1} + \frac{1}{p_2} - \frac{1}{p_1+p_2} = 10 + \frac{10}{9} - 1 = \frac{91}{9} \approx 10.11$$

- Per $N = 3, \mathbb{E}[X_{(\frac{1}{3},\frac{1}{3},\frac{1}{3})}] = 5.5$

  - ◦ Si $p_1 = \frac{1}{5}, p_2 = \frac{2}{5}, p_3 = \frac{2}{5}$, aleshores:

    $$\mathbb{E}[X_{(\frac{1}{5},\frac{2}{5},\frac{2}{5})}] = \frac{1}{p_1} + \frac{1}{p_2} + \frac{1}{p_3} - \frac{1}{p_1+p_2} - \frac{1}{p_1+p_3} - \frac{1}{p_2+p_3}$$
    $$+ \frac{1}{p_1+p_2+p_3} = 5 + \frac{5}{2} + \frac{5}{2} - \frac{5}{3} - \frac{5}{3} - \frac{5}{4} + 1 = \frac{77}{12} \approx 6.41$$

  - ◦ Si $p_1 = \frac{1}{6}, p_2 = \frac{2}{6}, p_3 = \frac{3}{6}$, aleshores:

    $$\mathbb{E}[X_{(\frac{1}{6},\frac{2}{6},\frac{3}{6})}] = \frac{1}{p_1} + \frac{1}{p_2} + \frac{1}{p_3} - \frac{1}{p_1+p_2} - \frac{1}{p_1+p_3} - \frac{1}{p_2+p_3}$$
    $$+ \frac{1}{p_1+p_2+p_3} = 6 + \frac{6}{2} + \frac{6}{3} - \frac{6}{3} - \frac{6}{4} - \frac{6}{5} + 1 = \frac{73}{10} = 7.3$$

En els exemples podem veure com varia el temps d'espera per completar una col·lecció segons les probabilitats d'obtindre cada cromo. En el primer exemple de cada apartat veiem el temps d'espera si un dels cromos té la meitat de probabilitats d'aparèixer que qualsevol altre.

A les figures 4 i 5, podem veure com varia el temps d'espera per completar una col·lecció en funció de les probabilitats d'obtindre un dels cromos de la col·lecció, considerant que la resta tenen les mateixes probabilitats de sortir.

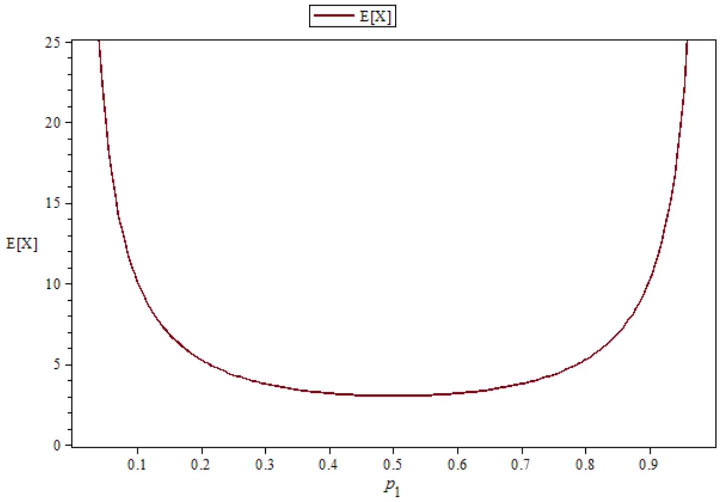

Figura 4: $\mathbb{E}[X]$ en funció de p_1 per $N = 2$.

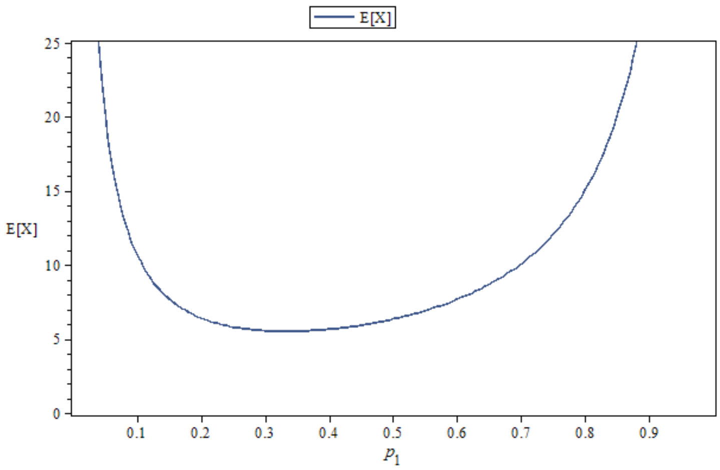

Figura 5: $\mathbb{E}[X]$ en funció de p_1 per $N = 3$ i $p_2 = p_3$.

Veiem que el mínim temps d'espera per completar una col·lecció es dóna en el cas que les probabilitats d'aconseguir cada cromo siguin $1/N$ en cada cas, és a dir, que tots els cromos tinguin les mateixes probabilitats d'aparèixer.

3.2 Càlcul usant integrals

En aquest apartat considerarem el mateix problema, calcular el temps d'espera per completar una col·lecció de N cromos que obtenim de manera unitària amb probabilitats desiguals. Usarem un plantejament alternatiu a l'apartat anterior i obtindrem una fórmula equivalent però més simplificada.

Designem per $n(t)$ el nombre de cromos que hem comprat a l'instant de temps t. Suposem que aquest procés es distribueix com un procés de Poisson de paràmetre $\lambda = 1$.

Considerem Y_i el temps transcorregut desde la compra $(i-1)$ fins la compra i. Aquestes variables segueixen una distribució exponencial de paràmetre $\lambda = 1$.

Igual que en l'apartat anterior, designem per X el nombre de cromos necessaris per completar la col·lecció.

Les variables X i Y_i són independents per tot i, ja que el temps que hi hagi entre una compra i la següent no afecta al nombre total de compres que hem de fer per completar la col·lecció, és a dir, haurem de comprar el mateix número de cromos tant si en comprem un cada dia com si en comprem un cada setmana, l'únic que variarà serà el temps que trigarem en completar la col·lecció.

Sigui Z_i el temps en el qual obtenim el cromo i per primera vegada. Designem per $Z = \max\{Z_1, ..., Z_N\}$ el temps en el qual ja tenim un cromo de cada, i per tant, la col·lecció completa.
Veiem que $Z = \sum_{i=1}^{X} Y_i$.

Proposició 3.2.1. Considerant les variables anteriors, aleshores $\mathbb{E}[Z] = \mathbb{E}[X]$.

Demostració.

$$\begin{aligned}
\mathbb{E}[Z] &= \mathbb{E}[\mathbb{E}[Z|X]] = \sum_k \mathbb{E}\left[\sum_{i=1}^{k} Y_i | X = k\right] \mathbb{P}(X = k) \\
&= \sum_k \mathbb{E}\left[\sum_{i=1}^{k} Y_i\right] \mathbb{P}(X = k) = \sum_k \sum_{i=1}^{k} \mathbb{E}[Y_i]\mathbb{P}(X = k) \\
\mathbb{E}[Z] &= \sum_k k\mathbb{P}(X = k) = \mathbb{E}[X]
\end{aligned}$$

□

Observació 3.2.2. Les variables aleatòries Z_i segueixen una distribució exponencial de paràmetre p_i, ja que,

$$\begin{aligned}
\mathbb{P}(Z_i > t) &= \sum_{k=0}^{\infty} \mathbb{P}(Z_i > t | n(t) = k)\mathbb{P}(n(t) = k) \\
&= \sum_{k=0}^{\infty} (1 - p_i)^k (e^{-t} \frac{t^k}{k!}) = e^{-t} \sum_{k=0}^{\infty} \frac{[(1 - p_i)t]^k}{k!} \\
&= e^{-t} e^{(1-p_i)t} = e^{-p_i t}.
\end{aligned}$$

Per tant,

$$\mathbb{P}(Z_i \leq t) = 1 - e^{-p_i t}.$$

Podem calcular $\mathbb{E}[Z]$ per donar resposta a la pregunta, quin és el nombre mitjà esperat de cromos que hem d'adquirir per completar la col·lecció.

Per definició, l'esperança de Z és:

$$\mathbb{E}[Z] = \int_0^{\infty} \mathbb{P}(Z > t)dt.$$

Considerant la funció de distribució de Z:

$$F_Z(t) = \mathbb{P}(Z \leq t) = \mathbb{P}(Z_1 \leq t, \ldots, Z_N \leq t) = \prod_{i=1}^{N} F_{Z_i}(t) = \prod_{i=1}^{N} (1 - e^{-p_i t}).$$

Podem reescriure l'esperança de Z per:

$$\mathbb{E}[Z] = \int_0^\infty \left(1 - \prod_{i=1}^{N}(1 - e^{-p_i t})\right) dt.$$

Hem obtingut una fórmula equivalent a l'anterior usant integrals.

Podem calcular el temps d'espera per completar una col·lecció de N cromos que arriben de manera unitària, amb probabilitats desiguals amb la fórmula:

$$\mathbb{E}[X] = \int_0^\infty \left(1 - \prod_{i=1}^{N}(1 - e^{-p_i t})\right) dt. \tag{5}$$

Exemple 3.2.3. Considerarem els exemples de l'apartat anterior calculats com a integrals:

$$\begin{aligned}
\mathbb{E}[X_{(\frac{1}{3},\frac{2}{3})}] &= \int_0^\infty \left(1 - \prod_{i=1}^{2}(1 - e^{-p_i t})\right) dt = \int_0^\infty \left(1 - (1 - e^{-\frac{1}{3}t})(1 - e^{-\frac{2}{3}t})\right) dt \\
&= \int_0^\infty \left(e^{-\frac{2}{3}t} + e^{-\frac{1}{3}t} - e^{-t}\right) dt = \left(-\frac{3}{2}e^{-\frac{2}{3}t} - 3e^{-\frac{1}{3}t} + e^{-t}\right)\Bigg|_{t=0}^{t=\infty} \\
\mathbb{E}[X_{(\frac{1}{3},\frac{2}{3})}] &= \left(\frac{3}{2} + 3 - 1\right) = \frac{7}{2} = 3.5
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[X_{(\frac{1}{5},\frac{2}{5},\frac{2}{5})}] &= \int_0^\infty \left(1 - \prod_{i=1}^{3}(1 - e^{-p_i t})\right) dt = \int_0^\infty \left(1 - (1 - e^{-\frac{1}{5}t})(1 - e^{-\frac{2}{5}t})^2\right) dt \\
&= \int_0^\infty \left(e^{-\frac{1}{5}t} + 2e^{-\frac{2}{5}t} - 2e^{-\frac{3}{5}t} - e^{-\frac{4}{5}t} + e^{-t}\right) dt \\
&= \left(-5e^{-\frac{1}{5}t} - 5e^{-\frac{2}{5}t} + \frac{10}{3}e^{-\frac{3}{5}t} + \frac{5}{4}e^{-\frac{4}{5}t} - e^{-t}\right)\Bigg|_{t=0}^{t=\infty} \\
\mathbb{E}[X_{(\frac{1}{5},\frac{2}{5},\frac{2}{5})}] &= (5 + 5 - \frac{10}{3} - \frac{5}{4} + 1) = \frac{77}{12} \approx 6.41
\end{aligned}$$

Veiem que, tot i que els càlculs són totalment diferents, els resultats obtinguts són els mateixos.

De fet, podem verificar que la fórmula 4 és equivalent a la fórmula 5 usant la igualtat següent:

$$\int_0^\infty e^{-pt}dt = -\frac{e^{-pt}}{p}\Bigg|_{t=0}^{t=\infty} = \frac{1}{p}$$

Reescribim la fórmula 4 obtinguda mitjançant "Maximums-Minimums":

$$\begin{aligned}
\mathbb{E}[X] &= \sum_i \int_0^\infty e^{-p_i t}dt - \sum_{i<j}\int_0^\infty e^{-(p_i+p_j)t}dt + \sum_{i<j<k}\int_0^\infty e^{-(p_i+p_j+p_k)t}dt - \dots \\
&\quad \dots + (-1)^{(N+1)}\int_0^\infty e^{-(p_1+\dots+p_N)t}dt \\
&= \int_0^\infty \left(\sum_i e^{-p_i t} - \sum_{i<j} e^{-(p_i+p_j)t} + \dots + (-1)^{(N+1)} e^{-(p_1+\dots+p_N)t}\right)dt
\end{aligned}$$

Usant que:

$$\prod_{i=1}^N (1-e^{-p_i t}) = 1 - \sum_i e^{-p_i t} + \sum_{i<j} e^{-(p_i+p_j)t} + \dots + (-1)^N e^{-(p_1+\dots+p_N)t}$$

Obtenim:

$$\mathbb{E}[X] = \int_0^\infty \left(1 - \prod_{i=1}^N (1-e^{-p_i t})\right)dt$$

Per tant hem comprovat que les dues fórmules són equivalents per calcular el temps d'espera de completar una col·lecció de N cromos que arriben de manera unitària i amb probabilitats desiguals.

3.3 Relació dels casos

Fins ara hem obtingut dues fórmules per calcular el temps d'espera per completar una col·lecció de N cromos, segons si les probabilitats d'obtindre cada cromo són iguals o diferents.

En el primer cas hem obtingut la fórmula:

$$F(N) \equiv \mathbb{E}[X] = N \sum_{i=1}^{N} \frac{1}{i}$$

Per al segon tenim que:

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{i} \left(\frac{1}{p_i}\right) - \sum_{i<j} \left(\frac{1}{p_i + p_j}\right) + \sum_{i<j<k} \left(\frac{1}{p_i + p_j + p_k}\right) - \dots \\
&\quad \dots + (-1)^{(N+1)} \left(\frac{1}{p_1 + \dots + p_N}\right)
\end{aligned}$$

En aquest apartat provarem que si $p_i = \frac{1}{N}$, $\forall i$, aleshores la fórmula obtinguda en el segon cas es redueix a l'anterior.

Primer de tot simplificarem la segona fórmula, considernat que $p_i = \frac{1}{N}, \forall i$. Obtenim:

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{i} \left(\frac{1}{\frac{1}{N}}\right) - \sum_{i<j} \left(\frac{1}{\frac{1}{N} + \frac{1}{N}}\right) + \sum_{i<j<k} \left(\frac{1}{\frac{1}{N} + \frac{1}{N} + \frac{1}{N}}\right) - \dots \\
&\quad \dots + (-1)^{(N+1)} \left(\frac{1}{\frac{1}{N} + \dots + \frac{1}{N}}\right) \\
\mathbb{E}[X] &= \sum_{i} N - \sum_{i<j} \frac{N}{2} + \sum_{i<j<k} \frac{N}{3} - \dots + (-1)^{(N+1)} \frac{N}{N} \\
\mathbb{E}[X] &= N \binom{N}{1} - \frac{N}{2} \binom{N}{2} + \frac{N}{3} \binom{N}{3} - \dots + \binom{N}{N} (-1)^{(N+1)} \\
\mathbb{E}[X] &= N \left(\binom{N}{1} - \frac{1}{2} \binom{N}{2} + \frac{1}{3} \binom{N}{3} - \dots + \binom{1}{N} (-1)^{(N+1)} \right) \\
\mathbb{E}[X] &= N \left(\sum_{i=1}^{N} \frac{1}{i} \binom{N}{i} (-1)^{(i+1)} \right)
\end{aligned}$$

Definim ara:

$$G(N) \equiv \mathbb{E}[X] = N \left(\sum_{i=1}^{N} \frac{1}{i} \binom{N}{i} (-1)^{(i+1)} \right)$$

Proposició 3.3.1. Si $p_i = \frac{1}{N}$, $\forall i$, aleshores $F(N) = G(N)$ per tot N.

És equivalent verificar que:

$$f(N) \equiv \sum_{i=1}^{N} \frac{1}{i} = \sum_{i=1}^{N} \frac{1}{i} \binom{N}{i} (-1)^{(i+1)} \equiv g(N)$$

Ho demostrarem per inducció ja que es tracta d'una proposició sobre els nombres naturals.

Les dues observacions següents ens resultaran molt útils per demostrar la proposició anterior.

Observació 3.3.2.

$$\binom{N}{i} = \frac{N!}{i!(N-i)!} = \frac{(N+1)!}{i!(N+1-i)!} \cdot \frac{(N+1-i)}{(N+1)} = \binom{N+1}{i} \cdot \frac{N+1-i}{N+1}$$

És a dir:

$$\binom{N}{i} = \binom{N+1}{i} [1 - \frac{i}{N+1}]$$

Observació 3.3.3.

$$0 = (1-1)^{N+1} = \sum_{i=0}^{N+1} \binom{N+1}{i} (-1)^i = 1 + \sum_{i=1}^{N} \binom{N+1}{i} (-1)^i + (-1)^{N+1}$$

Per tant,

$$0 = 1 - \sum_{i=1}^{N} \binom{N+1}{i} (-1)^{i+1} + (-1)^{N+1}$$

Aleshores:

$$\sum_{i=1}^{N} \binom{N+1}{i} (-1)^{i+1} = 1 + (-1)^{N+1}$$

Demostració. $f(N) = g(N)$
Primer de tot verifiquem que la proposició és vàlida per $N = 1$. És trivial veure que $f(1) = g(1) = 1$.
Suposem que la proposició és vàlida per N, és a dir, $f(N) = g(N)$ i la demostrarem per $N + 1$.

$$f(N+1) \;=\; \sum_{i=1}^{N+1}\frac{1}{i} = \sum_{i=1}^{N}\frac{1}{i} + \frac{1}{N+1} = f(N) + \frac{1}{N+1}$$

Per la hipòtesi d'inducció tenim que $f(N) = g(N)$, per tant,

$$f(N+1) \;=\; g(N) + \frac{1}{N+1} = \sum_{i=1}^{N}\frac{1}{i}\binom{N}{i}(-1)^{(i+1)} + \frac{1}{N+1}$$

Per la observació 3.3.2 podem canviar el nombre combinatori,

$$\begin{aligned}
f(N+1) &= \sum_{i=1}^{N}\frac{1}{i}\binom{N+1}{i}[1-\frac{i}{N+1}](-1)^{(i+1)} + \frac{1}{N+1} = \\
&= \sum_{i=1}^{N}\frac{1}{i}\binom{N+1}{i}(-1)^{(i+1)} - \sum_{i=1}^{N}\frac{1}{i}\binom{N+1}{i}\frac{i}{N+1}(-1)^{(i+1)} + \frac{1}{N+1} = \\
&= \sum_{i=1}^{N}\frac{1}{i}\binom{N+1}{i}(-1)^{(i+1)} - \frac{1}{N+1}\sum_{i=1}^{N}\binom{N+1}{i}(-1)^{(i+1)} + \frac{1}{N+1}
\end{aligned}$$

Per la observació 3.3.3 tenim:

$$\begin{aligned}
f(N+1) &= \sum_{i=1}^{N}\frac{1}{i}\binom{N+1}{i}(-1)^{(i+1)} - \frac{1}{N+1}[1+(-1)^{N+1}] + \frac{1}{N+1} = \\
&= \sum_{i=1}^{N}\frac{1}{i}\binom{N+1}{i}(-1)^{(i+1)} - \frac{1}{N+1} - \frac{(-1)^{N+1}}{N+1} + \frac{1}{N+1} = \\
&= \sum_{i=1}^{N}\frac{1}{i}\binom{N+1}{i}(-1)^{(i+1)} + \frac{(-1)^{N+2}}{N+1} = \\
&= \sum_{i=1}^{N+1}\frac{1}{i}\binom{N+1}{i}(-1)^{(i+1)}
\end{aligned}$$

Hem vist que:

$$f(N+1) = g(N+1)$$

Per tant, hem demostrat, per inducció sobre els nombres naturals, que $f(N) = g(N)$ per tot $N \in \mathbb{N}$, i per tant que $F(N) = G(N)$.

$\square$

4 Un problema relacionat

Considerem el problema següent:

Donada una distribució discreta uniforme, determinar la cardinalitat mínima d'una mostra aleatòria d'aquesta distribució, en la qual es puguin observar un nombre prefixat m d'elements diferents.

Si considerem m, el nombre prefixat d'elements diferents, igual a la cardinalitat de la distribució donada estem reduint el problema al cas del col·leccionista de cromos que hem resolt fins ara, ja que, volem obtindre tots els elements de la distribució.

En canvi, considerant el problema amb $m < N$ podem donar resposta a una altre pregunta del col·leccionista: quants cromos haig de comprar per obtindren m diferents?

Considerem el cas d'una col·lecció de N cromos que apareixen amb les mateixes probabilitats.
Sigui $S = \{1, \ldots, N\}$ el suport de la distribució discreta uniforme. Tenim que l'element i, apareixerà a la mostra amb una probabilitat $p_i = \frac{1}{N}$ per tot $i \in \{1, \ldots, N\}$, de fet, coincideix amb la probabilitat que té el cromo i d'aparèixer.

Considerem una sèrie de variables aleatòries:
Sigui X_i el nombre de cromos que comprem per passar de tenir-ne $(i-1)$ diferents a tenir-ne i diferents.
Designem per $X_N(m) = X_1 + \ldots + X_m$ el nombre de cromos que hem de comprar per aconseguir m elements diferents.
Les variables aleatòries X_i, per $i \in \{1, \ldots, N\}$, segueixen una llei geomètrica de parametre $\frac{N-i+1}{N}$.

Per tant, per la linearitat de l'esperança:

$$
\begin{aligned}
\mathbb{E}[X_N(m)] &= \mathbb{E}[X_1 + \ldots + X_m] \\
\mathbb{E}[X_N(m)] &= 1 + \frac{N}{N-1} + \frac{N}{N-2} + \ldots + \frac{N}{N-m+2} + \frac{N}{N-m+1} \\
\mathbb{E}[X_N(m)] &= N \sum_{i=1}^{m} \frac{1}{N-i+1} = N \sum_{i=N-m+1}^{N} \frac{1}{i}
\end{aligned}
$$

Veiem que per $m = N$ obtenim la solució del temps d'espera per completar una col·lecció de N cromos.

Exemple 4.0.4. Suposem que tenim una col·lecció de 10 cromos diferents. Pel que hem vist en aquest apartat, el temps d'espera per aconseguir 9 cromos diferents és:

$$\mathbb{E}[X_{10}(9)] \quad = \quad N \sum_{i=N-m+1}^{N} \frac{1}{i} = 10 \sum_{i=2}^{10} \frac{1}{i} = \frac{4861}{2520} \approx 19.29$$

Com haviem vist anteriorment, el temps d'espera per completar una col·lecció de 10 cromos és 29.29, per tant, un cop aconseguits 9 cromos diferents, el temps d'espera per completar la col·lecció són 10 cromos. De fet, restant a la fórmula 1 la que acabem d'obtindre, aconseguim una nova fórmula per calcular el temps d'espera per completar una col·lecció quan ja tenim m cromos diferents.

Definim per $\mathbb{E}[T_N(m)]$ el temps d'espera per completar una col·lecció de N cromos, sabent que ja tenim m cromos diferents. En aquest cas:

$$\begin{aligned} \mathbb{E}[T_N(m)] &= \mathbb{E}[X] - \mathbb{E}[X_N(m)] = N\left(\sum_{i=1}^{N} \frac{1}{i} - \sum_{i=N-m+1}^{N} \frac{1}{i}\right) \\ \mathbb{E}[T_N(m)] &= N \sum_{i=1}^{N-m} \frac{1}{i} \end{aligned}$$

En l'exemple anterior,

$$\mathbb{E}[T_{10}(9)] = 10 \sum_{i=1}^{1} \frac{1}{i} = 10$$

A continuació veurem que obtindre la primera meitat dels cromos d'una col·lecció és molt més ràpid que obtindre la segona meitat. Les fórmules obtingudes en aquest apartat ens permeten calcular el temps d'espera intermitjos d'una col·lecció.

A la taula 3 hi ha diferents exemples del temps d'espera per aconseguir la primera meitat dels cromos d'una col·lecció. També hi consten el temps d'espera per acabar la col·lecció suposant que ja tenim la primera meitat dels cromos.
Observem que el temps que triguem en aconseguir la segona meitat dels cromos és molt superior a la primera, ja que al principi no tenim gaires cromos a

N	m	$\mathbb{E}[\boldsymbol{X_N(m)}]$	$\mathbb{E}[\boldsymbol{T_N(m)}]$
6	3	3.7	11
10	5	6.46	22.83
30	15	20.3	99.55
50	25	34.16	190.8
100	50	68.82	449.92
200	100	138.13	1037.47
500	250	346.07	3050.34

Taula 3: Temps d'espera per completar mitja col·lecció i per acabar-la.

la col·lecció, per tant, la probabilitat que no tinguem un cromo que adquirim serà superior que en el cas que ja tinguem mitja col·lecció, que aleshores, la probabilitat que no tinguem un cromo que adquirim serà inferior.

Finalment, hem vist que a mesura que aconseguim un cromo nou a la col·lecció, el temps d'espera per aconseguir el següent que no estigui repetit augmenta.

5 Arribada múltiple

Considerem el problema del col·leccionista de cromos, calcular el temps d'espera per completar una col·lecció, en el cas que els cromos arribin en paquets de varies unitats.

Suposem que els cromos s'adquireixen en grups de mida constant g, amb $1 < g < N$. Suposem també que a cada grup de cromos no hi pot haver cap cromo repetit. D'aquesta manera tenim que el nombre total de grups de cromos diferents és $\binom{N}{g}$.

Identifiquem els elements que conté cada grup de cromos amb els vectors $(a_1, \cdots, a_g) \in \{1, \cdots, N\}^g$, amb $a_i < a_{i+1}$, per tot $i = 1, \cdots, g-1$.

Usem l'ordre lexicogràfic per ordenar els diferents grups de cromos.

Def. 5.0.5. Diem que el grup de cromos $(a_1, \cdots, a_g) < (b_1, \cdots, b_g)$ si existeix $i \in \{1, \cdots, g-1\}$ tal que $a_s = b_s$ per $s < i$, i $a_i < b_i$.

Designem per q_i, $i \in \{1, \cdots, \binom{N}{g}\}$ la probabilitat d'obtindre l'i-èssim grup de cromos, ordenats segons l'ordre lexicogràfic que acabem de definir.

Separarem el problema de l'arribada múltiple segons si tots els grups de cromos tenen les mateixes probabilitats de ser adquirits o no. És a dir, segons els valors de q_i.

5.1 Probabilitats iguals

En aquest cas, cada grup de cromos té la mateixa probabilitat de ser adquirit, per tant, $q_i = \frac{1}{\binom{N}{g}}$, per tot $i = \{1, \ldots, \binom{N}{g}\}$.

Sigui V_i el nombre de grups de cromos a comprar per obtindre una unitat del cromo i. Aquestes variables aleatòries segueixen una llei geomètrica de paràmetre $1 - \frac{\binom{N-1}{g}}{\binom{N}{g}}$, ja que $\frac{\binom{N-1}{g}}{\binom{N}{g}}$ és la probabilitat de no obtindre el cromo i.

Les variables aleatòries $\min(V_i, V_j)$ segueixen una llei geomètrica de paràmetre $1 - \frac{\binom{N-2}{g}}{\binom{N}{g}}$, ja que $\frac{\binom{N-2}{g}}{\binom{N}{g}}$ és la probabilitat de no obtindre ni el cromo i ni el cromo j.

Finalment tenim que les variables aleatòries $\min(V_{i_1}, \cdots, V_{i_{N-g}})$ segueixen una llei geomètrica de paràmetre $1 - \frac{1}{\binom{N}{g}}$.

Les variables aleatòries $\min(V_{i_1}, \cdots, V_{i_k})$ amb $k > N - g + 1$ són iguals a la variable aleatòria constant 1.

Per resoldre el problema utilitzem la fórmula de "Maximums-Minimums", vista a l'apartat 3.1, que ens permet calcular el nombre esperat de grups de cromos a adquirir per aconseguir un cromo de cada i així completar la col·lecció.

Definim per $V = \max(V_1, \cdots, V_N)$ el nombre de grups de cromos a adquirir per completar la col·lecció.

$$\begin{aligned}\mathbb{E}[\max(V_1, \cdots, V_N)] = \sum_{1\le i\le N} \mathbb{E}[V_i] - \sum_{1\le i<j\le N} \mathbb{E}[\min(V_i, V_j)] + \cdots \\ +(-1)^{N-g+1} \sum_{1\le i_1<i_2<\cdots<i_{N-g}\le N} \mathbb{E}[\min(V_{i_1}, \ldots, V_{i_{N-g+1}}) + \\ +(-1)^{N-g+2} \sum_{1\le i_1<i_2<\cdots<i_{N-g+1}\le N} 1 + \cdots + (-1)^{N+1}\end{aligned}$$

$$\begin{aligned}\mathbb{E}[V] = \binom{N}{1} \frac{1}{1 - \frac{\binom{N-1}{g}}{\binom{N}{g}}} - \binom{N}{2} \frac{1}{1 - \frac{\binom{N-2}{g}}{\binom{N}{g}}} + \cdots \\ \cdots + (-1)^{N-g+1} \binom{N}{N-g} \frac{1}{1 - \frac{1}{\binom{N}{g}}} + \\ + \sum_{1\le k\le g} (-1)^{N-g+k+1} \binom{N}{N-g+k}\end{aligned}$$

Amb aquesta fórmula podem calcular el temps d'espera per completar una col·lecció de N cromos que arriben en grups de mida constant g i tots els grups aparèixen amb les mateixes probabilitats.

Exemple 5.1.1. A la taula 4 veiem com varia el nombre de grups de cromos a comprar per completar una col·lecció en funció del nombre de cromos a la col·lecció i el nombre de cromos a cada grup. A la taula 5 veiem que el nombre de cromos a comprar, sigui en grups o individualment, no varia excessivament, a no ser que el tamany dels grups de cromos sigui significatiu respecte al tamany de la col·lecció.

N	$\boldsymbol{g=1}$	$\boldsymbol{g=2}$	$\boldsymbol{g=3}$	$\boldsymbol{g=4}$	$\boldsymbol{g=5}$
5	11.42	5.33	3.22	2.25	1
10	29.29	14.12	9.05	6.49	4.94
30	119.85	59.16	38.92	28.8	22.71
50	224.96	111.59	73.8	54.9	43.55
100	518.74	258.32	171.51	128.1	102.05
200	1175.61	586.58	390.23	292.09	233.16
500	3396.41	1696.76	1130.2	846.92	676.96

Taula 4: Grups a adquirir en funció de N i g.

N	$\boldsymbol{g=1}$	$\boldsymbol{g=2}$	$\boldsymbol{g=3}$	$\boldsymbol{g=4}$	$\boldsymbol{g=5}$
5	11.42	10.65	9.67	9	5
10	29.29	28.25	27.14	25.95	24.67
30	119.85	118.32	116.76	115.18	113.57
50	224.96	223.19	221.4	219.6	217.77
100	518.74	516.63	514.52	512.39	510.26
200	1175.61	1173.16	1170.71	1168.25	1165.79
500	3396.41	3393.51	3390.61	3387.71	3384.8

Taula 5: Cromos a adquirir en grups en funció de N i g.

5.2 Probabilitats desiguals

En aquest apartat considerem que cada grup de cromos té una probabilitat diferent de ser adquirit. Calcularem el nombre de grups de cromos que hem de comprar per completar una col·lecció de N cromos.

Igual que en l'apartat anterior, suposem que els grups estan ordenats seguint l'ordre lexicogràfic, i denotem per q_i la probabilitat d'obtindre l'i-èssim grup de cromos.

Def. 5.2.1. Definim la probabilitat de comprar un grup de cromos que no conté cap unitat dels cromos $i_1, \cdots, i_k$, per $q(i_1, \cdots, i_k)$, per tot $k \in \{1, \cdots, N-g\}$.

$$q(i_1) = \sum_{i=\binom{N-1}{g-1}+1}^{\binom{N}{g}} q_i$$

$$q(i_1, i_2) = \sum_{i=\binom{N-1}{g-1}+\binom{N-2}{g-1}+1}^{\binom{N}{g}} q_i$$

Finalment obtenim que:

$$q(i_1, i_2, \cdots, i_k) = \sum_{i=\binom{N-1}{g-1}+\cdots+\binom{N-k}{g-1}+1}^{\binom{N}{g}} q_i \quad \text{si } k \leq N-g$$
$$q(i_1, i_2, \cdots, i_k) = 0 \quad \text{si } k > N-g$$

Considerem, igual que en l'apartat anterior, les variables aleatòries V_i, on per cada i, V_i és el nombre de grups de cromos a comprar per obtindre una unitat del cromo i. Aquestes variables segueixen una llei geomètrica de parametre $1 - q(i)$.

Les variables $\min(V_i, V_j)$ segueixen una llei geomètrica de paràmetre $1 - q(i, j)$.

Les variables $\min(V_i, \cdots, V_k)$ segueixen una llei geomètrica de paràmetre $q(1, \cdots, N-g)$ si $k \leq N-g+1$, i en canvi són iguals a la variable aleatòria constant 1 si $k > N-g+1$.

Aleshores, aplicant la fórmula de "Maximums-Minimums", obtenim:

$$\begin{aligned}
\mathbb{E}[V] &= \sum_{1\le i\le N} \frac{1}{1-q(i)} - \sum_{1\le i<j\le N} \frac{1}{1-q(i,j)} + \cdots \\
&\quad \ldots + (-1)^{N-g+1} \sum_{1\le i_1<i_2<\cdots<i_{N-g}\le N} \frac{1}{1-q(i_1,\cdots,i_{N-g})} + \\
&\quad + \sum_{1\le k\le g} (-1)^{N-g+k+1} \binom{N}{N-g+k}
\end{aligned}$$

Hem obtingut una nova fórmula per calcular el nombre de grups de cromos a comprar per completar una col·lecció de N cromos diferents, que adquirim en grups de mida constant g, i suposant que cada grup de cromos té una probabilitat q_i d'aparèixer.

Aquesta fórmula es redueix a l'obtinguda a l'apartat anterior si suposem que $q_i = \frac{1}{\binom{N}{g}}$, és a dir, si tots els cromos tenen les mateixes probabilitats d'aparèixer.

Exemple 5.2.2. Suposem que volem calcular el temps d'espera per completar una col·lecció de 4 cromos que arriben en grups de mida 2. Aleshores tenim $\binom{4}{2} = 6$ grups de cromos diferents.

Els numerem seguint l'ordre lexicogràfic definit a l'apartat anterior.

$$(1,2) < (1,3) < (1,4) < (2,3) < (2,4) < (3,4)$$

Les probabilitats d'obtindre cada grup de cromos són $q_1, q_2, \cdots, q_6$, respectivament.

En aquest cas $(N-g) = 2$, per tant només hem de calcular $q(i_1, \cdots, i_k)$ per k fins a 2. Tenim que:

$$q(1) = q_4 + q_5 + q_6 \;,\; q(2) = q_2 + q_3 + q_6 \;,\; q(3) = q_1 + q_3 + q_5 \;,$$
$$q(4) = q_1 + q_2 + q_4 \;,\; q(1,2) = q_6 \;,\; q(1,3) = q_5 \;,\; \ldots \;,\; q(3,4) = q_1.$$

Aleshores podem calcular el nombre de grups de cromos a adquirir per completar la col·lecció usant la fórmula:

$$\mathbb{E}[V] = \sum_{i=1}^{4} \frac{1}{1-q(i)} - \sum_{1\le i<j\le 4} \frac{1}{1-q(i,j)} + \sum_{k=1}^{4} (-1)^{3+k} \binom{4}{2+k}$$

Veiem uns casos particulars:

- $\mathbb{E}[V_{(\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6})}] = \frac{19}{5} = 3.8$
- $\mathbb{E}[V_{(\frac{1}{11},\frac{2}{11},\frac{2}{11},\frac{2}{11},\frac{2}{11},\frac{2}{11})}] = \frac{347}{90} \approx 3.86$
- $\mathbb{E}[V_{(\frac{45}{50},\frac{1}{50},\frac{1}{50},\frac{1}{50},\frac{1}{50},\frac{1}{50})}] = \frac{161387}{6909} \approx 23.36$

En el primer cas hem suposat que les probabilitats d'obtindre cada grups de cromos són les mateixes. De fet, aquest exemple també es pot calcular amb la fórmula obtinguda a l'apartat anterior, ja que es tracta del cas particular que les probabilitats de cada grup de cromos siguin iguals.

En el segon exemple, hem suposat que el primer grup de cromos apareix la meitat de vegades que qualsevol altre, i podem comprobar que aquest fet no afecta significativament al nombre de grups de cromos que hem de comprar per completar la col·lecció, ja que, podem aconseguir els cromos del primer grup amb els altres grups de cromos on surtin.

Per últim, hem suposat que el primer grup de cromos apareix un noranta per cent de les vegades. En aquest cas sí que veiem que el nombre de grups de cromos a adquirir per completar la col·lecció es dispara, ja que, les possibilitats d'obtindre els cromos que no són al primer grup han disminuit notablement.

A la figura 6 podem veure el nombre de grups de cromos a comprar per completar una col·lecció de 4 cromos en funció de les probabilitats d'obtindre un dels grups de cromos. Hem suposat que els altres grups de cromos tenen les mateixes probabilitats de ser adquirits.

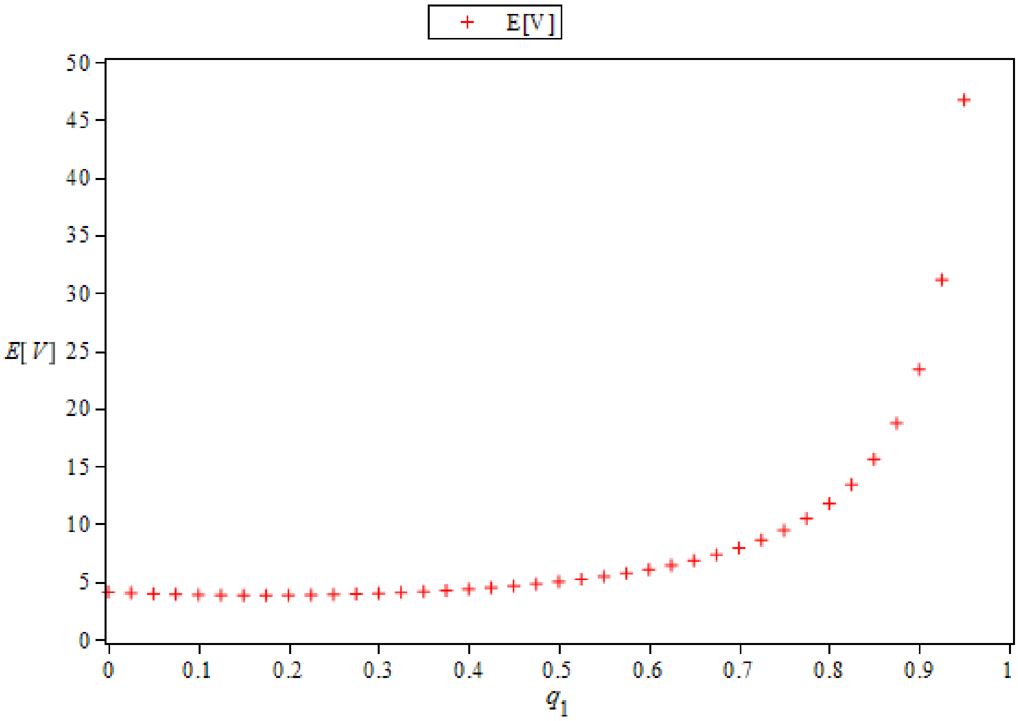

Figura 6: $\mathbb{E}[V]$ en funció de q_1 per $N = 4$ i $q_2 = \ldots = q_6$.

Veiem que, el mínim de grups de cromos a comprar per completar la col·lecció es dóna en el cas que les probabilitats d'obtindre cada grup siguin les mateixes. També veiem que el fet que es disminueixi la probabilitat d'obtindre un sol grup no afecta significativament al nombre de grups de cromos a comprar, però en canvi, si un grup de cromos apareix molt més sovint que la resta, fa que el temps d'espera es dispari cap a l'infinit.

6 Conclusions

En aquest treball ens hem endinsat en el món del col·leccionisme. Hem suposat el cas particular de col·leccions de cromos, però, de fet, es poden aplicar aquests resultats a qualsevol col·lecció d'elements diferents que obtenim de manera aleatòria.

A partir dels resultats obtinguts, concluim que el temps d'espera per completar una col·lecció de cromos varia en funció de diversos paràmetres.

Primerament hem considerat el fet que, sovint, obtenim els elements de la col·lecció en grups de mida constant. Hem comprovat les diferències que hi ha entre obtindre els elements en grups o obtindre'ls de forma individual. Podem concloure que si el nombre d'elements a cada grup no es significatiu respecte a la mida de la col·lecció, el fet que obtinguem els cromos d'un en un o en grups no afecta significativament al nombre final esperat de cromos que hem de comprar per completar la col·lecció.

Una altra consideració és si tots els cromos apareixen amb les mateixes probabilitats o no. Concluim que les probabilitats d'obtindre cada cromo afecten de manera directa sobre el temps d'espera per completar la col·lecció, ja que, com hem vist, la probabilitat d'obtindre un sol cromo pot fer que el temps d'espera per completar la col·lecció augmenti de manera notable. També hem verificat que el temps d'espera per completar una col·lecció és minimitza si les probabilitats d'obtindre cada cromo són iguals.

Finalment, concloure que aquest tema encara es pot ampliar, considerant noves suposicions, com ara el fet de cooperar amb altres col·leccionistes per completar la col·lecció.

Agraïments

A Xavier Bardina, per la proposta i tutorització d'aquest treball, i posar-me en contacte amb Marco Ferrante, un dels autors citats a al bibliografía, que em va transmitir la seva curiositat sobre el tema de les col·leccions de cromos.

Referències

[1] S.Ross, *A first course in probability*, 9th Edition, Pearson, 2012.

[2] M.Ferrante, N.Frigo, *A note in the coupon-collector's problem with multiple arrivals and the random sampling.*

[3] M.Ferrante, N.Frigo, *On the expected number of different records in a random sample.*

[4] A.Boneh, M.Hofri, *The coupon-col·lector problem revisited - a survey of engineering problems and computational methods.* Computer Science Technical Reports. Paper 807 (1989).

[5] F.Hillier, G.Lieberman, *Introducción a la investigacion de operaciones*, McGraw Hill, 2001.

[6] B.James, K.James, *An approximation for the coupon collectors's problem*, University of Minnesota, 2012.

[7] V.Doumas, G.Papanicolaou, *Some new aspects for the random coupon collector's problem*, ALEA, 2014.

[8] P.Flajolet, D.Gardy, L.Thimonier, *Birthday paradox, coupon collectors, caching algorithms and self-organizing search*, Discrete Applicated Mathematics, 1992.

[9] L.Holst, *On birthday, collectors, occupancy and other classical urn problems*, International Statistics, 1986.